IMM 363
January 1968

Courant Institute of
Mathematical Sciences

Minimax Linear Predictor under Lipschitz' Type Conditions for the Regression Function

Kei Takeuchi

Prepared under Contract Nonr-285(38)
with the Office of Naval Research

Distribution of this document is unlimited.

New York University

NR 042-206

IMM 363
January 1968

New York University

Courant Institute of Mathematical Sciences

MINIMAX LINEAR PREDICTOR UNDER LIPSCHITZ' TYPE CONDITIONS FOR THE REGRESSION FUNCTION

Kei Takeuchi

This report represents results obtained at the Courant Institute of Mathematical Sciences, New York University, with the Office of Naval Research, Contract Nonr-285(38).

ABSTRACT

Suppose that $Y_i = f(x_i) + U_i$, $i = 1,\ldots,n$, and $E(U_i) = 0$, $E(U_i^2) = \sigma^2$, $E(U_i U_j) = 0$, $i \neq j$.

We want to estimate $f(x_0)$ by a linear function $\hat{f}(x_0) = \sum a_i Y_i$. Those coefficients a_i which minimize

$$\sup_f E(\hat{f}(x_0) - f(x_0))^2 \quad \text{for some class of } f$$

are sought. In two cases where

(a) $\left|\dfrac{f(x_i) - f(x_j)}{x_i - x_j}\right| \leq c\sigma$,

(b) $x_i = i$, $x_0 = 0$

$|\Delta^{h+1} f(k)| \leq c\sigma$ where $\Delta f(k) = f(k+1) - f(k)$,

the minimax solution is obtained, and it is shown that the solution coincides with weight least squares with weight

$$w_i = \max\,(\lambda - \mu|x_i - x_j|,\ 0) .$$

§0. Introduction.

Regression analysis is the most (or at least one of the most) popular and most often used techniques in various fields of statistical data analysis. In some cases, however, regression analysis is very dangerous, and sometimes gives awkward results. Such dangers, which are inherent in regression techniques, are well known, at least well perceived by experienced applied statisticians. But theoretical analysis of such a situation that yields some pitfalls to the careless application of regression analysis is far from satisfactory. Though well trained statisticians can evade such a danger by their good judgment, there is no formal well established technique that may be applied. The purpose of this paper is to derive some method to treat one such difficulty, i.e. the problem of the functional form of the regression.

Suppose that we have a quantity or response Y, which is influenced by some quantity or explanatory variable x, and we assume a relation, as

$$Y = f(x) + U ,$$

where $f(x)$ denotes the influence of x on Y and U is a random variable, which may be the error of measurement, or random shock or something similar. We assume that n pairs of observations (x_i, Y_i), $i = 1,\ldots,n$, are given,

and for each pair

$$Y_i = f(x_i) + U_i \; ; \qquad i = 1,\ldots,n$$

and $U_1,\ldots,U_n$ are distributed independently and with mean 0 and variance σ^2. Such assumptions on the distribution of U may well be the source of trouble in many practical situations, but we shall not discuss the problem here, and simply take it for granted that the above assumptions are true.

Our problem is to estimate the form of the function f(x). But without any assumption on f, we can go no further, and it is usually assumed that f(x) is, or nearly is, a linear function, and it is expressed as

$$Y_i = \alpha + \beta x_i + U_i \; , \qquad i=1,\ldots,n,$$

where α and β are unknown parameters and are estimated by least squares.

In many cases a linear function gives a reasonably good approximation to the function f(x), or random error is so large that any indication of nonlinearity of f(x) does not show up. But in other cases, and not in rare cases, the data significantly contradict the linearity assumption of the function f(x). Then the quadratic function will be used, and if the quadratic function is still unsatisfactory, cubic, or polynomials of higher degree will be used. And if sufficiently high degrees are allowed, the data will be

fitted very closely by a polynomial of the type

$$\hat{f}(x) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_p x^p \ .$$

Since any continuous function can be approximated to any precision by a polynomial, the above procedure may seem plausible; for, in practical situations, it can be assumed without any doubt that $f(x)$ is continuous, and smooth enough. But it also happens in practical situations that a polynomial of higher degree fits the data very nicely, but gives riduculous values when extrapolated.

The simplest example is the case when x is the time t. We have a time series data for some period, for $t = 1, 2, \ldots, T$, and we would like to predict the value of Y for the next period, i.e. for $t = T+1$. In this case, if no other relevant information is available, we assume a model of the type

$$Y = \beta_0 + \beta_1 t + \ldots + \beta_p t^p + u_t \ ,$$

and estimating $\beta_0 \ldots \beta_p$ by least squares, and substituting $t = T+1$, we have a predicted value of Y for the next period. And it often happens that when the data is very long, though a simple function such as a linear or quadratic function fails to be fitted to the data, using a high degree polynomial, we have riduculous values for the next period. Consequently, it may seem intiutively plausible to use only the latter part

of the data and to fit a simple function such as one that is linear or quadratic in order to predict the near future. Or more sophisticatedly, we can use a system of weight $w_1\ w_2\ \ldots\ w_T$, which is increasing with t (i.e. decreasing as the data tends to the distant past), and fitting a function by weighted least squares using those weights. And it will be shown actually that such a weight system, i.e. a system for which $w_t = 0$ for $t \leq t_0$ and $w_t = a - bt$, for $t > t_0$, is optimum in some minimax sense.

We shall deal with the problem from the viewpoint of prediction, or estimation of the value of the function at the specific value of $x = x_0$. And if $\hat{f}(x_0)$ is a predictor or estimator, we shall use as a criterion

$$E(\hat{f}(x_0) - f(x_0))^2$$

i.e. the mean square error. And we assume that nothing definite is known about the functional form of f(x) but that it satisfies some regularity conditions. And we shall seek a predictor which minimizes the supremum of mean square error with respect to the class of functions satisfying those conditions.

§1. Minimax Linear Predictor under the First Order Lipschitz condition.

First we shall consider the simplest case. Let (x_i, Y_i), $i = 1,2,\ldots,n$, be n pairs of data, and assume that

$$Y_i = f(x_i) + U_i$$

and

$$E(U_i) = 0 \ , \ i = 1,2,\ldots,n; \qquad E(U_i^2) = \sigma^2, \ i = 1,2,\ldots,n:$$

$$E(U_i \ U_j) = 0 \ , \ i \neq j.$$

We want to predict the value of f at $x = x_0$, i.e. $f(x_0)$, by a linear combination of Y_i,

$$\hat{f}(x_0) = a_1 Y_1 + \ldots + a_n Y_n = \sum a_k Y_k \ .$$

Then the mean square error of $\hat{f}(x_0)$ is given by

$$E\{(\hat{f}(x_0) - f(x_0))^2\} = \{\sum a_k f(x_k) - f(x_0)\}^2 + \sum a_k^2 \sigma^2 \ .$$

We shall consider the following assumption.

Assumption 1. For any $x_i \ x_j$, $x_i \neq x_j$, there is a constant c which satisfies

$$\left| \frac{f(x_i) - f(x_j)}{x_i - x_j} \right| \leq c\sigma \ . \tag{1.1}$$

And we are to minimize the supremum of mean square error with respect to all functions which satisfy the above

assumption.

In order that the supremum be finite, we must have

$\sum a_k = 1$.

We assume that $x_1 \leq x_2 \leq \dots \leq x_j \leq x_0 \leq x_{j+1} \leq \dots \leq x_n$

and define

$$d_k = x_{k+1} - x_k \qquad \text{for } k = 1,2,\dots,j-1$$

$$d_j = x_0 - x_j$$

$$d_{j+1} = x_{j+1} - x_0$$

$$d_k = x_k - x_{k-1} \qquad \text{for } k = j+2,\dots,n,$$

and also define

$$\Delta_k = f(x_{k+1}) - f(x_k) \qquad \text{for } k = 1,2,\dots,j-1$$

$$\Delta_j = f(x_0) - f(x_j)$$

$$\Delta_{j+1} = f(x_{j+1}) - f(x_0)$$

$$\Delta_k = f(x_k) - f(x_{k-1}) \qquad \text{for } k = j+2,\dots,n.$$

Then

$$f(x_k) = f(x_0) - \Delta_j - \dots - \Delta_k \text{ for } k \leq j$$

$$f(x_k) = f(x_0) + \Delta_{j+1} + \dots + \Delta_k \text{ for } k \leq j+1 .$$

Hence, if $\sum a_k = 1$,

$$\sum a_k f(x_k) - f(x\ ) = - \sum_{k=1}^{j} a_k \sum_{i=k}^{j} \Delta_i + \sum_{k=j+1}^{n} a_k \sum_{i=j+1}^{k} \Delta_i$$

$$= - \sum_{i=1}^{j} \Delta_i \sum_{k=1}^{i} a_k + \sum_{i=j+1}^{n} \Delta_i \sum_{k=1}^{n} a_k.$$

Assumption 1 is equivalent to $|\Delta_k| \leq cd_k\sigma$, $k = 1,\ldots,n$, so that under Assumption 1,

$$\sup \{\sum a_k f(x_k) - f(x_0)\}^2 = c^2\sigma^2\{\sum_{i=1}^{j} d_i \sum_{k=1}^{i} a_k + \sum_{i=j+1}^{n} d_i \sum_{k=1}^{n} a_k \}^2.$$

Without loss of generality we can assume that $\sigma = 1$, and the problem is to minimize

$$c^2\{\sum_{i=1}^{n} d_i|b_i|\}^2 + \sum a_k^2 \quad \text{where} \quad b_i = \sum_{k=1}^{i} a_k \quad \text{or} \quad \sum_{k=i}^{n} a_k \tag{1.2}$$

under the condition that $\sum a_k = 1$.

To minimize (1.2), it will be shown that we must have $a_k \geq 0$ for all k.

Suppose that we have one negative element $a_\ell < 0$, $1 \leq \ell \leq j$. It is easily shown that b_j and b_{j+1} must be nonnegative. Hence either $b_\ell = \sum_{i=1}^{\ell} a_i \geq 0$ or $b_\ell < 0$, and there is one pair $\ell < h$ such that $a_\ell < 0$, $a_{\ell+1} = \ldots = a_{h-1} = 0$, $a_h > 0$. Then define a new system of coefficients a_k' by

$$a_\ell' = 0 \text{ and } a_k' = b_\ell a_k / b_{\ell-1}, \qquad \text{for } k < \ell;$$
$$= a_k, \qquad \text{otherwise};$$

for the first case, and

$$a'_\ell = a_\ell + c, \quad a'_h = a_h - c, \qquad \text{where } c = \min(a_\ell, a_h));$$

$$a'_k = a_k \qquad\qquad \text{otherwise;}$$

for the second case. And in either case we have

$$|a'_k| \leq |a_k| \quad \text{and} \quad |b'_i| \leq |b_i| \qquad\qquad \text{for all } k,i$$

and

$$|a'_\ell| < |a_\ell| \ .$$

Hence we have

$$c^2 \{\sum_{i=1}^{n} d_i |b'_i|\}^2 + \sum a'^2_k < c^2\{\sum d_i \ |b_i|\}^2 + \sum a_k^2 \ .$$

Thus the problem is reduced to minimizing

$$c^2 \{\sum_{i=1}^{j} d_i \sum_{k=1}^{i} a_k + \sum_{i=j+1}^{n} d_i \sum_{k=1}^{n} a_k\}^2 + \sum a_k^2$$

$$= c^2 \{\sum_{k=1}^{j} (\sum_{i=k}^{j} d_i) a_k + \sum_{k=j+1}^{n} (\sum_{i=j+1}^{k} d_i) a_k\}^2 + \sum a_k^2 \qquad (1.3)$$

under the condition that $\sum a_k = 1$ and $a_k \geq 0$.

Denoting the Lagrangian multiplier by 2λ and differentiating, we have for the condition of minimum,

$$\begin{aligned} c^2\delta_k M + a_k - \lambda &= 0 \qquad \text{if} \quad a_k > 0 \\ &\geq 0 \qquad \text{if} \quad a_k = 0 \end{aligned} \qquad (1.4)$$

where $\delta_k = \sum_{i=k}^{j} d_i = |x_0 - x_k| \quad \text{for } i \leq j$

$\qquad = \sum_{i=j+1}^{k} d_i = |x_0 - x_k| \quad \text{for } i \geq j+1 \ ;$

and $M = \sum \delta_k a_k$.

Or, in other words,

$$a_k = \max(\lambda - c^2 M\delta_k, 0)$$
$$M = \sum \delta_k a_k , \qquad (1.5)$$
$$\sum a_k = 1$$

There is a unique set of solutions for a_k and λ which satisfies (1.5), which is calculated as follows. Let the number of the sample be reordered according to the size of δ, i.e. let $0 \leq \delta_1 < \delta_2 < \ldots < \delta_n$. Then $a_k > 0$ for $k \leq m$ for some m and $a_k = 0$ for $k \geq m+1$, and from (1.5)

$$c^2 M\, \delta_m < \lambda \leq c^2 M\, \delta_{m+1}$$

and

$$M = \sum_{k=1}^{m} \delta_k \lambda - c^2 M \sum_{k=1}^{m} \delta_k^2 .$$

Hence

$$\sum_{k=1}^{m} \delta_k(\delta_m - \delta_k) < \frac{1}{c^2} \leq \sum_{k=1}^{m+1} \delta_k(\delta_{m+1} - \delta_k) = \sum_{k=1}^{m+1} \delta_k(\delta_{m+1} - \delta_k).$$

Define $R_m = \sum_{k=1}^{m} \delta_k(\delta_m - \delta_k)$, then R_m is monotone increasing and m is determined from

$$R_m < \frac{1}{c^2} \leq R_{m+1} .$$

Once m is determined, λ and M are determined easily from (1.5).

There may be some complexity if some of the δ's are equal, but the argument above applies also in such cases, without any serious modification.

Thus, it is established that the following theorem holds true.

Theorem 1. Under Assumption 1, the coefficients of linear minimax predictor in the sense that it minimizes the supremum of mean square error, are given uniquely by (1.5).

Moreover, the supremum of mean square error for the minimax predictor is given by

$$\lambda = \frac{1 + c^2 \sum_{k=1}^{m} \delta_k^2}{m(1 + c^2 \sum_{k=1}^{m} (\delta_k - \bar{\delta})^2} \qquad \text{where } \bar{\delta} = \frac{1}{m} \sum \delta_k .$$

The latter half of the theorem is obtained from what follows. The maximum mean square error for given a_k is equal to

$$c^2 M^2 + \sum a_k^2 .$$

But, for the minimax predictor,

$$c^2 M \, \delta_k a_k + a_k^2 - \delta a_n = 0 \qquad \text{for all } k$$

$$c^2 M \sum \delta_k a_k + \sum a_k^2 = c^2 M^2 + \sum a_k^2 = \lambda \sum a_k = \lambda .$$

And λ is obtained from the equations:

$$\sum a_k = m\lambda - c^2 M \sum \delta_k = 1$$

$$\sum \delta_k a_k = (\sum \delta_k)\lambda - c^2 M \sum \delta_k^2 = M.$$

§2. Minimax predictor under higher order conditions.

Next we shall consider a more complicated situation, that is we shall consider a higher order Lipschitz condition. On the other hand, we assume simply that x_i are placed at equal distances, i.e. we shall assume that $x_i = i$, $i = 1,\dots,n$, and that $x_0 = 0$.

We define a difference operator Δ by

$$\Delta f(x) = f(x+1) - f(x)$$

and power Δ^j of Δ by

$$\Delta^j(f(x)) = \Delta(\Delta^{j-1} f(x+1) - \Delta^{j-1} f(x)), \quad j = 2,3,\dots .$$

We assume that

Assumption 2. $\quad |\Delta^{h+1} f(x)| \le c\sigma,$

and we shall obtain a minimax linear predictor under this assumption. For simplicity we put $\sigma = 1$ as before.

Let $\hat{f}(0) = \sum a_k Y_k$ be a linear predictor. Then

$$E(\hat{f}(0) - f(0))^2 = (\sum a_k f(k) - f(0))^2 + \sum a_k^2,$$

and $f(k)$ can be expressed as

$$f(k) = f(0) + {}_kc_1 d_1 + {}_kc_2 d_2 + \cdots + {}_kc_k d_k$$

$$+ \sum_{j=0}^{k-h-1} {}_{k-j-1}c_k \Delta^{h+1} f(j) , \quad k = 1,2,\ldots, \tag{2.1}$$

where $d_i = \Delta^i f(0)$, ${}_kc_i$ is the binomial coefficient with the definition that ${}_kc_i = 0$ if $1 \leq k < h$.

If no restrictions are imposed on $f(0)$ and $d_1 \ldots d_k$, we must have

$$\sum_{k=1}^{n} a_k = 1 , \quad \sum_{k=1}^{n} {}_kc_j a_k = 0 , \quad j = 1,2,\ldots,k, \tag{2.2}$$

in order that $\sup E(\hat{f}(0) - f(0))^2 < \infty$.

And under the assumption above,

$$\sup E(\hat{f}(0) - f(0)) = c^2 (\sum_{j=0}^{n-h-1} |b_j|)^2 + \sum a_k^2 \tag{2.3}$$

where $b_j = \sum_{k=j+h+1}^{n} {}_{k-j-1}c_h a_k$.

Thus it is necessary to minimize (2.3) under the condition (2.2).

Let the Lagrangian form Φ be

$$\Phi(a_1 \ldots a_n) = c^2 (\sum_{j=0}^{n-h-1} | \sum_{k=j+h+1}^{n} {}_{k-j-1}c_h a_k |)^2 + \sum a_k^2$$

$$- 2 \lambda_0 \sum a_k - 2\lambda_1 \sum {}_kc_1 a_k \cdots$$

$$- 2 \lambda_h \sum {}_kc_h a_h .$$

Since ϕ is strictly convex in its arguments, a local minimum gives the global minimum, and if for suitable $\lambda_0 \ldots \lambda_h$, the local minimum satisfies the condition (2.2) it gives the solution of our problem. And a necessary and sufficient condition for local minimality is given by

$$\left(\frac{\partial\phi}{\partial a_k}\right)_+ \geq 0\ , \quad \left(\frac{\partial\phi}{\partial a_k}\right)_- \leq 0\ , \qquad k = 1, 2, \ldots, n,$$

where $(\partial\phi/\partial a_k)_+$ and $(\partial\phi/\partial a_k)_-$ are the right and the left derivatives respectively, which are given by

$$\frac{1}{2}\left(\frac{\partial\phi}{\partial a_k}\right)_+ = c^2 M \sum_{j=0}^{k-h-1} {}_{k-j-1}c_h \ \mathrm{sgn}\ (b_j+0) + a_k - \lambda_0 - \lambda_1\ {}_k c_1$$

$$- \ldots - \lambda_h\ {}_k c_h\ ;$$

$$\frac{1}{2}\left(\frac{\partial\phi}{\partial a_k}\right)_- = c^2 M \sum_{j=0}^{k-h-1} {}_{k-j-1}c_h \ \mathrm{sgn}\ (b_j+0) + a_k - \lambda_0 - \lambda_1\ {}_k c_1$$

$$- \ldots - \lambda_n\ {}_k c_h\ ,$$

where $\quad b_j = \sum_{k=j+h+1}^{n} {}_{k-j-1}c_h a_k\ , \quad M = \sum_{j=0}^{n-h-1} |b_j|$ and

$$\begin{aligned} \mathrm{sgn}\ (b+0) &= 1 \quad &&\text{if} \quad b \geq 0 \\ &= -1 && \phantom{\text{if}} \quad b < 0 \\ \mathrm{sgn}\ (b-0) &= 1 &&\text{if} \quad b > 0 \\ &\ -1 && \phantom{\text{if}} \quad b \leq 0\ . \end{aligned}$$

We shall conjecture that the solution will have the form

$$a_k = 0 \quad \text{for } k = m+1 \ldots n \quad \text{and } a_m \neq 0 \text{ for some } m$$
$$\text{and } b_j \text{ are of the same sign for all } j. \tag{2.4}$$

If this conjecture were correct, for $k \leq m$ we have

$$a_k = \beta_0 + \beta_1 k + \ldots + \beta_{h+1} k^{h+1} \tag{2.5}$$

since for $k \leq m$,

$$\left(\frac{\partial \phi}{\partial a_k}\right)_+ = \left(\frac{\partial \phi}{\partial a_k}\right)_{-1} = 0 \ .$$

We shall prove that this conjecture is actually true, which will be done in several steps.

First we shall define the following terminology.

<u>Definition</u>. For a sequence $a_1 \ldots a_m$, we shall say that it changes sign at a_i, if

$$a_i a_{i+1} < 0 \ , \text{ or for some } j \quad a_{i+1} = \ldots = a_{i+j-1} = 0$$
$$\text{and} \qquad a_i a_{i+j} < 0 \ .$$

Then we have

(1) If for $a_1 \ldots a_m$, $\sum_{k=1}^{m} {}_k c_j \ a_k = 0$ for $j = 1,\ldots,h$, $a_1 \ldots a_m$ changes sign at least h times provided that $a_k \not\equiv 0$.

<u>Proof</u>. The condition implies that

$$\sum_{k=1}^{m} {}_k c^j \ a_k = 0 \ , \qquad k = 1,\ldots,h \ .$$

Assume that $a_1,\ldots,a_m$ changes sign at $a_{i_1},\ldots,a_{i_r}$, and define a polynomial

$$g(t) = t(i_1 + \frac{1}{2} - t) \ldots (i_r + \frac{1}{2} - t) .$$

If $r \leq h-1$, we have from the assumption

$$\sum g(k)\, a_k = 0 .$$

But by the definition of g, for all k we must have

$$\begin{aligned} g(k)a_k &\geq 0 \quad \text{if the first non-zero element} > 0 \\ &\leq 0 \quad \text{if it is} \leq 0, \end{aligned}$$

which implies that $a_k \equiv 0$. Q.E.D.

(2) For a_k satisfying the condition of the solution as well as (2.4) and (2.5) we must have

$$a_k = \beta(\alpha_1 - k) \ldots (\alpha_{h+1} - k) \tag{2.6}$$

where $1 \leq \alpha_1 < \ldots < \alpha_h < m < \alpha_{h+1}$ and $\beta > 0$, and $\alpha_{h+1} \leq m + 1$ if $m < n$.

Proof. Since $b_{n-h-1} = {}_mC_h a_m$,

$$\operatorname{sgn} |b_j| \equiv \operatorname{sgn} |a_m|$$

if all b_j have the same sign. Hence

$$\frac{1}{2}\left(\frac{\partial\phi}{\partial a_m}\right)_+ = \frac{1}{2}\left(\frac{\partial\phi}{\partial a_m}\right)_- = c^2 M\, {}_mC_{h+1} \operatorname{sgn} |a_m| + a_m - \lambda_0 - \lambda_1\, {}_mC_1 - \cdots$$
$$- \lambda_h\, {}_mC_h = 0$$

implies that the signs of β_{h+1} in (2.5) and a_m are opposite.

By (1) it is shown that the equation

$$\beta_0 + \beta_1 x + \dots + \beta_{k+1} x^h = 0$$

must have at least h distinct roots in the interval (1,m), and the above implies that it has one which is larger than m.

And if $\beta_{h+1} > 0$, then $a_{m+1} = 0$ for the solution implies

$$\left(\frac{\partial\phi}{\partial a_{m+1}}\right)_+ = (-1)^{h+1} \beta(\alpha_1 - (m+1)) \dots (\alpha_{h+1} - (m+1)) \geq 0 ;$$

hence $m+1 \geq \alpha_{h+1}$.

Similarly for $\beta_{h+1} < 0$, $(\partial\phi/\partial a_{m+1})_- \leq 0$ implies $m+1 \geq \alpha_{h+1}$.

And

$$\sum_k (\alpha_1 - k) \dots (\alpha_h - k) a_k = \sum \alpha_1 \dots \alpha_h a_k = \alpha_1 \dots \alpha_h$$

$$= \sum \beta(\alpha_1 - k)^2 \dots (\alpha_h - k)^2 (\alpha_{h+1} - k),$$

implies that $\beta > 0$. Q.E.D.

(3) If $a_k = \beta(\alpha_1 - k), \dots, (\alpha_h - k)(\alpha_{h+1} - k)$ for $k \leq m$

$= 0$ for $k \geq m+1$

and $\beta > 0$, $1 < \alpha_1 < \dots < \alpha_h < m < \alpha_{h+1}$,

and it holds that

$$\sum {}_k c_j \, a_k = 0 , \quad j = 1, \dots, h.$$

Then $(-1)^h b_j = (-1)^h \sum_{k=j+h+1}^{m} {}_{k-j-1}C_h a_k > 0$ for $j = 0, 1, \ldots, m-h-1$.
Define a function

$$g_{h,j}(x) = (x-j-1)\ldots(x-j-h)/h! \ , \qquad x \geq j+h,$$
$$\qquad = 0 \qquad\qquad\qquad\qquad , \qquad x \leq j+h.$$

We shall show that for any polynomial $h(x)$ of degree h, a sequence $\{c_k\}$ defined by $c_k = h(k) - g_{k,j}(k)$, $k = 1, 2, \ldots$ cannot change the sign more than $h+1$ times.

If $h = 1$, this is obvious.

For $h \geq 2$, define a new sequence $\{c_k\}$ by

$$c_k' = \Delta c_k = c_{k+1} - c_k \ , \qquad k = 1, 2, \ldots,$$

then $c_k' = h'(k) - g_{h-1,j}(k)$, where $h'(k)$ is a polynomial of degree $h-1$, and if $\{c_k\}$ changes sign more than $h+1$ times, $\{c_k'\}$ must change sign more than h times. Hence, by induction, the proposition is proved.

Define a polynomial $h^*(x) = \gamma_1 x + \ldots + \gamma_h x^h$ such that

$$h^*(\alpha_j) = g_{h,j}(\alpha_j) \quad \text{for } j = 1, \ldots, h.$$

Then the equation $h^*(x) - g_{h,j}(x) = 0$ has $h+1$ distinct roots, $0, \alpha_1, \ldots, \alpha_h$, and no other roots than that. Hence $\gamma_h > 0$, and since $1 < \alpha_1$, $(-1)^h(h^*(1) - g_{h,j}(1)) < 0$. Thus,

$$(-1)^h(h^*(k) - g_{h,j}(k))a_k \leq 0 \quad \text{for all } k = 1, 2, \ldots, m,$$

the equality being satisfied only when $a_k = 0$.

Hence

$$0 > \sum (-1)^h (h^*(k) - g_{h,j}(k))a_k$$

$$= (-1)^h \sum h^*(k)a_k - (-1)^h \sum g_{h,j}(k)a_k$$

$$= - (-1)^h b_j ,$$

the equality being omitted since $a_m \neq 0$. Q.E.D.

(4) For any $c^2 > 0$, there is an α^* and a polynomial

$$\psi(k) = \gamma_0 + \gamma_1 k + \ldots + \gamma_h k^h$$

such that

$$\sum_{k=1}^{m} (\alpha^* - k)\psi(k) = 1$$

$$\sum_{k=1}^{m} {}_m c_j(\alpha^* - k)\psi(k) = 0 , \qquad j = 1,2,\ldots,h,$$

$$\sum_{k=1}^{m} {}_m c_{h+1}(\alpha^* - k)\psi(k) = (-1)^h \gamma_h/c^2 ,$$

where $m = \min(n, [\alpha^*])$.

<u>Proof</u>. First we shall show that for fixed $\alpha \geq m+1$, there is a polynomial

$$\bar{\psi}_\alpha(k) = \bar{\gamma}_0 + \bar{\gamma}_1 k + \ldots + k^h ,$$

for which

$$\sum_{k=1}^{m} {}_kc_j\ (\alpha - k)\bar{\psi}_\alpha(k) = 0\ , \quad j = 1,2,\ldots,h, \tag{2.6}$$

where $m = \min(n, [\alpha])$, if $n \geq m+1$.

We shall express $\bar{\psi}_\alpha(k)$ in the form

$$\bar{\psi}_\alpha(k) = \lambda_0' + \lambda_1'\ {}_kc_1 + \ldots + \lambda_{h-1}'\ {}_kc_{h-1} + {}_kc_h\ .$$

Then putting $w_k = \sqrt{\alpha - k}$, equation (2.6) can be expressed as

$$\sum_{j'=0}^{h-1} \Big(\sum_{k=1}^{m} (w_k\ {}_kc_j)(w_k\ {}_kc_{j'})\Big)\lambda_j' = - \sum_{k=1}^{m} (w_k\ {}_kc_j)(w_k\ {}_kc_h),$$

$$j' = 0\ldots h-1, \tag{2.7}$$

Since as functions of k, $w_k\ {}_kc_j$ are linearly independent if $m \geq h+1$, the coefficient matrix of λ' in (2.7) is non-singular, and there is a unique solution in $\lambda_1' \ldots \lambda_h'$.

Next we shall show that

$$\ell(\alpha) = \sum_{k=1}^{m} {}_kc_{h+1}\ \bar{\psi}_\alpha(k)(\alpha - k)$$

is a monotone increasing function of α.

Suppose that for $\alpha_1 < \alpha_2$, $m_{\alpha_1} = m_{\alpha_2} = m \geq h+1$. Then $\ell(\alpha_2) - \ell(\alpha_1) = \sum_{k=1}^{m} {}_kc_{h+1}\ \psi^*(k)$,

where $\psi^*(k)$ is a polynomial of degree h, and we have

$$\sum_{k=1}^{m} {}_kc_j\ \psi^*(k) = 0\ , \qquad \text{for } j = 1,\ldots,h.$$

By (1), $\{\psi^*(k)\}$, $k = 1,2,\ldots,$ changes sign h times, and

hence it has the form

$$\psi^*(k) = \beta'(\alpha_1' - k)\ldots(\alpha_h' - k)$$

where $1 < \alpha_1' < \ldots \alpha_h' < m$. Substituting $k = \alpha_1$

$$\psi^*(\alpha_1) = \bar{\psi}_{\alpha_2}(\alpha_1)\ (\alpha_2-\alpha_1)\ .$$

Since $\bar{\psi}_{\alpha_2}(k) = 0$ has h roots all smaller than m, $\bar{\psi}_{\alpha_2}(\alpha_1) > 0$, hence

$$\psi^*(\alpha_1) > 0\ , \quad (-1)^h\,\beta' > 0\ .$$

Since

$$\sum k^j\,\psi^*(k) = 0\ , \qquad j = 1,2,\ldots,h,$$

$$\begin{aligned}\sum k^{c_{h+1}}\,\psi^*(k) &= \sum k^{h+1}\,\psi^*(k)\\ &= \sum k(k-\alpha_1')\ldots(k-\alpha_h')\psi^*(k)\\ &= (-1)^h\beta' \sum k(k-\alpha_1')^2\ldots(k-\alpha_h')^2 > 0.\end{aligned}$$

Thus it is shown that $\ell(\alpha)$ is monotone increasing in each interval $m \leq \alpha \leq m+1$, where m is an integer, and since $\ell(h+1) = 0$, and $\ell(\alpha) \to 0$ as $\alpha \to \infty$, $\ell(\alpha)$ is monotone increasing from 0 to infinity, as α increases from $\alpha = h+1$. Hence there is a unique solution that $\ell(\alpha^*) = 1/c^2$. And it is also easily shown that $\sum (\alpha - k)\bar{\psi}^*_\alpha(k) \neq 0$, and by defining γ_h by $\gamma_h = (\sum(\alpha-k)\bar{\psi}_\alpha(k))^{-1}$,

and putting $\psi(k) = \gamma_h \bar{\psi}_\alpha = (k)$ we have the desired solution. Q.E.D.

Now we can prove the main theorem of this section.

<u>Theorem</u> 2. Let α^* and $\psi(k)$ be defined as in (4) above. Then

$$a_k = (\alpha^* - k)\psi(k) \quad \text{for } k \leq m ,$$
$$= 0 \quad \text{for } k \geq m+1,$$

gives the coefficients of the minimax linear predictor under Assumption 2.

<u>Proof</u>. From (1) and (3), for such a_k, $(-1)^h b_j = \sum {}_{k-j-1}c_h a_k$ > 0 for $j = 0,1,\ldots.$

Hence

$$\left(\frac{\partial \phi}{\partial a_k}\right)_+ = \left(\frac{\partial \phi}{\partial a_k}\right)_{-1} = 0 \quad \text{for } k = 1,\ldots,m.$$

And, if $m+1 \geq \alpha^*$,

$$\frac{1}{2}\left(\frac{\partial \phi}{\partial a_k}\right)_+ = -(\alpha^* - k)\psi(k) \geq 0 \quad \text{if } h \text{ is even}$$

$$\frac{1}{2}\left(\frac{\partial \phi}{\partial a_k}\right)_- = -(\alpha^* - k)\psi(k) \leq 0 \quad \text{if } h \text{ is odd},$$

for $k \geq m+1$.

Put for even h

$$-(\alpha^* - k)\psi(k) = \mu\, {}_k c_{h+1} + \lambda_0 + \lambda_1 c_1 + \cdots + \lambda_h\, {}_k c_h$$

$$= \mu\, {}_k c_{h+1} + \psi_1(k) .$$

Then $\mu > 0$ and $\psi_1(k)$ is a polynomial of degree h.

We shall show that if the sequence a_k, $k=1,2,\ldots,$ defined by $a_k = \mu_k \, c_{h+1} + \psi_1(k)$ where $\psi_1(k)$ is a polynomial of degree, h changes sign h+1 times, and suppose that a_m is the term at which the $h+1^{th}$ change occurs, then $\psi_1(k)$ is negative and monotone decreasing for $k \geq m$. The proof will be done by induction.

When $h = 0$, $\psi_1(k)$ being a constant it is obvious.

When $h \geq 1$, let a new sequence be defined by

$$a_k' = \Delta a_k = a_{k+1} - a_k = \mu_k c_h + \psi_1'(k) \ ,$$

and $\psi_1'(k)$ is a polynomial of degree h-1, and defined by

$$\psi_1'(k) = \psi(k+1) - \psi(k) \ .$$

Then $\{a_k'\}$ changes sign h times, and the h-th change occurs not after a_m'. Thus if the proposition were true for h-1, then

$$\psi_1'(k) < 0 \text{ for all } k \geq m.$$

And since $a_m = \mu \, {}_m c_h + \psi(m) < 0$, we have $\psi(m) < 0$ and $\psi(k) < 0$ for all $k \geq m$.

By this proposition with even h, it is shown that

$$\frac{1}{2} \left(\frac{\partial \phi}{\partial a_k}\right)_+ \Big|_{a_k=0} = \mu \, {}_m c_{h+1} - \mu \sum_{\ell=m+1}^{k} {}_\ell c_h + \lambda_0 + \lambda_1 \, {}_k c_1$$

$$+ \ldots + \lambda_k \, {}_k c_h$$

$$< \mu \; {}_m c_{h+1} + \psi_1(k)$$

$$= \mu \; {}_m c_{h+1} + \psi_1(m) + (\psi_1(k) - \psi_1(m))$$

$$= a_m + (\psi_1(k) - \psi_1(m)) < 0 \quad \text{for } k \geq m+1.$$

Quite similarly for odd h it is shown that

$$\frac{1}{2} \left(\frac{\partial \phi}{\partial a_k}\right)_+ \Big|_{a_k = 0} > 0 \quad \text{for } k \geq m+1.$$

Which complites the proof. Q.E.D.

It should be noted that the solution given by Theorem 2 above coincides with fitting polynomials of degree h by weighted least squares, with weight $w_k = \alpha^* - k$ for $k < \alpha^*$ and $w_k = 0$ for $k \geq \alpha^*$.

To see this, suppose that

$$\hat{f}(x) = \hat{\beta}_0 + \hat{\beta}_1 x + \ldots + \hat{\beta}_h x^h ,$$

be the weighted least squares estimator for $\hat{f}(x)$.

Then $f(0) = \beta_0$ will be a linear combination of Y of the form

$$\hat{\beta}_0 = \lambda_0 \sum_{k=1}^{n} w_k Y_k + \lambda \sum_{k=1}^{n} w_k \, k Y_k + \ldots + \lambda_h \sum_{k=1}^{n} w_k \, k^h \, Y_k$$

$$= \sum (\alpha^* - k)(\lambda_0 + \lambda_1 k + \ldots + \lambda_h k^h) Y_k$$

$$= \sum a_k Y_k \; .$$

And a_k must satisfy

$$\sum a_k = 1, \quad \sum {}_k c_j a_k = 0 , \quad j = 1,\ldots,h.$$

Thus a_k must be equal to those previously obtained.

Arithmetically diminishing weights may seem intuitively appealing, and in fact it was shown to be optimum in the above sense, such a procedure may be recommended in practical situations.

§3. Some Numerical Discussion.

We shall briefly discuss numerical properites of the results of the previous section.

If α^* defined in the previous section is equal to $m+1$, a_k is obtained from the relation that

$$a_k = \alpha_0 + \alpha_1 k + \alpha_2 k^2 + \ldots + \alpha_{h+1} k^{h+1} = \phi_h(k) \quad \text{for } k \leq m$$

and

$$\phi_h(m+1) = 0 ,$$

$$\sum_{k=1}^{m} \phi_h(1) = 1 ,$$

$$\sum_{k=1}^{m} {}_k c_j \; \phi_h(k) = 0 , \qquad j = 1,\ldots,h.$$

And it is shown after some algebraic calculations that

$$\phi_0(k) = \frac{2}{m(m+1)} (m+1-k)$$

$$\phi_1(k) = \frac{6}{m(m+1)(m-1)} (m+1-k)(m+1-2k)$$

$$\phi_2(k) = \frac{12}{m(m+1)(m-1)(m-2)} (m+1-k)(m^2+2m+2-5(m+1)k+5k^2) \tag{3.1}$$

$$\phi_3(k) = \frac{20}{m(m+1)(m-1)(m-2)(m-3)} (m+1-k)(m+1-2k)$$

$$\times (m^2+2m+6-7(m+1)k+7k^2).$$

And these functions give the coefficients of minimax predictors corresponding to the case when

c^2 = absolute value of the coefficient of the highest power in $\phi_h(k) = \sum {}_kC_{h+1}\ \phi(k)$

or when

$$c^2 = c_m^2 = \frac{K_h}{{}_{m+h+2}C_{2h+3}} \tag{3.2}$$

where $K_0 = 1$, $K_1 = 4$, $K_2 = 15$, $K_3 = 56$, etc.

And the supremum of the mean square error is simply equal to $\phi_h(0)$, which will be chosen as follows.

It was shown that

$$a_k = \phi_h(k) = \lambda_0 + \lambda_1\ {}_kC_1 + \ldots + \lambda_h\ {}_kC_h + \lambda_{h+1}\ {}_kC_{h+1}$$

and $\lambda_{h+1} = -c^2M$, $M = \sum {}_kC_{h+1}\ a_k$.

And the mean square error is given

$$c^2M^2 + \sum a_k^2 = c^2M \sum {}_kc_{h+1}a_k + \sum a_k^2$$

$$= \sum_k a_k(c^2M \; {}_kc_h + a_k)$$

$$= \sum_k a_k(\lambda_0 + \lambda_1 \; {}_kc_1 + \dots + \lambda_h \; {}_kc_h) = \lambda_0 = \phi_h(0) \; .$$

And from (3.1) it is obtained

$$\begin{aligned} \phi_0(0) &= \frac{2}{m} \\ \phi_1(0) &= \frac{6(m+1)}{m(m-1)} \\ \phi_2(0) &= \frac{12(m^2+2m+2)}{m(m-1)(m-2)} \\ \phi_3(0) &= \frac{20(m+1)(m^2+2m+6)}{m(m-1)(m-2)(m-3)} \end{aligned} \tag{3.3}$$

When c^2 is given, we should choose m such that

$$c_m^2 \geq c^2 > c_{m+1}^2$$

and $a_k = {}_h(k)$ may be obtained in the way indicated in the previous section.

We shall show some of the numerical values below.

c_m^2

h \ m	1	2	3	4	5	6	7	8	9	10
0	1	$\frac{1}{4}$	$\frac{1}{10}$	$\frac{1}{20}$	$\frac{1}{35}$	$\frac{1}{56}$	$\frac{1}{84}$	$\frac{1}{120}$	$\frac{1}{165}$	$\frac{1}{220}$
1	∞	4	$\frac{2}{3}$	$\frac{4}{21}$	$\frac{1}{14}$	$\frac{2}{63}$	$\frac{1}{63}$	$\frac{2}{231}$	$\frac{1}{198}$	$\frac{4}{1287}$
2		∞	15	$\frac{15}{8}$	$\frac{5}{12}$	$\frac{3}{8}$	$\frac{3}{22}$	$\frac{5}{88}$	$\frac{15}{572}$	$\frac{15}{1144}$
3			∞	56	$\frac{28}{5}$	$\frac{56}{55}$	$\frac{14}{55}$	$\frac{56}{715}$	$\frac{4}{143}$	$\frac{8}{715}$

$\phi_h(0)$

h \ m	1	2	3	4	5	6	7	8	9	10
0	2	1	$\frac{2}{2}$	$\frac{1}{2}$	$\frac{2}{5}$	$\frac{1}{3}$	$\frac{2}{7}$	$\frac{1}{4}$	$\frac{2}{9}$	$\frac{1}{5}$
1	∞	9	4	$\frac{5}{2}$	$\frac{9}{5}$	$\frac{7}{5}$	$\frac{8}{7}$	$\frac{27}{28}$	$\frac{5}{6}$	$\frac{11}{15}$
2		∞	34	13	$\frac{37}{5}$	5	$\frac{26}{7}$	$\frac{41}{14}$	$\frac{101}{42}$	$\frac{61}{30}$
3			∞	125	41	21	$\frac{42}{7}$	$\frac{129}{14}$	$\frac{125}{18}$	$\frac{11}{2}$

The value of the supremum of mean square error for intermediate values of c^2 is troublesome to calculate, but rough idea of it will be obtained by interpolation. We shall denote by e^2 the maximum mean square error, and have an approximate formula about the relation between e^2 and c^2. When h = 0

$$c_m^2 = \frac{6}{m(m+1)(m+2)} \doteqdot \frac{6}{(m+1)^3},$$

and

$$e^2 = \frac{2}{m} \tag{3.4}$$

for integral values of m, hence we can have approximately,

$$e^2 \doteq \frac{2}{(6/c^2)^{1/3}-1} \doteq (\tfrac{2}{3})^{1/3}\ c^{2/3} \quad \text{for small c.}$$

Similarly for h = 1

$$e^2 \doteq \frac{6}{(480/c^2)^{1/5}-3} \doteq (\tfrac{81}{5})^{1/5}\ c^{2/5}\ ,$$

and for h = 2,

$$e^2 \doteq \frac{12}{(15\times 7!/c^2)^{1/7}-6} \doteq (\frac{2^{10}3^4}{175})^{1/7}\ c^{2/7}; \tag{3.4$'$}$$

for h = 3,

$$e^2 \doteq \frac{20}{(56 \times 9!/c^2)^{1/9}-10} \doteq (\frac{10^8}{3^4 7^2})^{1/9}\ c^{2/9}\ .$$

Generally, for small c, we have

$$e^2 \doteq \eta_h\ c^{2/(2h+1)}\ ,$$

where η_h is a constant, and seems to be slowly increasing with h.

In practical situations we may not usually be quite sure about h, and also about c^2, so that the above discussions may give us some indications about the choice of h. And

usually it will be safe to assume rather large value for c^2, and if we take c^2 not to be very small, m must be small, and e^2 increases rapidly with h. As a rough conclusion, we say that it is safer to take rather small h, and not too large m.

As an example, suppose that the true regression function is something like

$$f(k) = \alpha \beta^k$$

and we assume that $\alpha > 0$ and $\beta < 1$. (It should be remembered that in our discussion the order of the time sequence is reversed for the next period is denoted by $k = 0$.)
Then

$$|\Delta^h f(k)| \leq \alpha(1-\beta)^h , \qquad h = 1,2,\dots .$$

If we are quite sure that $\alpha \leq 10$, $\beta \geq 0.8$, then we have the following values for c^2,

$$c^2 = 100 \times (0.2)^{2h}.$$

And we have actually

$$\begin{aligned} m &= 1 , & &\text{for } h = 0 \\ m &= 2 , & &h = 1 \\ m &= 7 , & &h = 2 \\ m &= 9 , & &h = 3 . \end{aligned}$$

And approximate values for e^2 are given by,

$$\begin{aligned} e^2 &= 5\,, &&\text{for } h = 0 \\ e^2 &= 4\,, && h = 1 \\ e^2 &= 3.9, && h = 2 \\ e^2 &= 9\,, && h = 3\,. \end{aligned}$$

Hence, in this case the best predictor will fit a quadratic function using 7 values. And in this case $c^2 = 0.16$ may be replaced by $c_u^2 = 3/22 = 0.14$, and, using this, we have from (3.5) the weight system, from (3.1),

$$\frac{21}{14}\,,\ \frac{3}{14}\,,\ -\frac{5}{14}\,,\ -\frac{6}{14}\,,\ -\frac{3}{14}\,,\ \frac{1}{14}\,,\ \frac{3}{14}\,.$$

And if actually $f(k) = 10 \times 0.8^k$, the mean square error of the predictor with the above weight is shown to be equal to 3.56.

Usually, c^2 does not decrease so rapidly as in the exponential case, so that we may have stronger reason to use a rather small value of h.

§4. Application to the Smoothing of Time Series.

The results thus far obtained can be applied to interpolation or smoothing of time series, i.e. in case x_0 is equal to some x_i.

We shall discuss in this section only the simplest case. That is, we assume that $x_i = i$, $i = 1,\ldots,T$, and

the Assumption 1 in §1 is taken into consideration.

In this case, the whole discussion in §1 can be applied, and we have the following results.

Suppose that $x_0 = t$, $1 \leq t \leq T$. Define τ by

$$\tau = \min\,(t-1,\ T-t)\ . \tag{4.1}$$

Then using the notation of §1, we have

$$\delta_1 = 0,\ \delta_2 = \delta_3 = 1,\ \ldots,\ \delta_{2\tau} = \delta_{2\tau+1} = \tau,\ \delta_{2\tau+2} = \tau+1,$$
$$\ldots,\ \delta_T = T - \tau - 1\ .$$

We define R_k by

$$R_1 = 0,\ R_{2k} = R_{2k+1} = \sum_{i=1}^{2k+1} \delta_i(\delta_{2k+1} - \delta_i) \qquad \text{for } k \leq \tau\ ,$$

$$R_k' = \sum_{i=1}^{k'} \delta_i(\delta_k - \delta_i)\ , \qquad \text{for } k' \geq \tau\ .$$

We determine m to be the largest integer for which R_m is smaller than $1/c^2$. Then we have exactly m of a_k's with positive values. The rest is a matter of simple numerical computation.

In this way, we can have a consistent way of interpolation for all $t = 1,\ldots,T$, provided that c^2 is the same for all the cases, and we can also extrapolate consistently.

The following are numerical examples.

Assume that $c^2 = 1/2$; then it is shown that if $\tau \geq 1$,

$$a_{t-1} = \frac{1}{4}, \quad a_t = \frac{1}{2}, \quad a_{t+1} = \frac{1}{4},$$

and for $t = 1$, we have with the same value for c^2,

$$a_1 = \frac{7}{12}, \quad a_2 = \frac{1}{3}, \quad a_3 = \frac{1}{12}, \qquad \text{for } t = 1,$$

$$a_T = \frac{7}{12}, \quad a_{T-1} = \frac{1}{3}, \quad a_{T-2} = \frac{1}{12}, \qquad \text{for } t = T.$$

Similarly for $c^2 = 1/8$,

$$a_{t-2} = \frac{1}{9}, \quad a_{t-1} = \frac{2}{9}, \quad a_t = \frac{1}{3}, \quad a_{t+1} = \frac{2}{9}, \quad a_{t+1} = \frac{1}{9}, \quad \text{for } t \geq 2,$$

$$a_1 = \frac{16}{66}, \quad a_2 = \frac{23}{66}, \quad a_3 = \frac{16}{66}, \quad a_4 = \frac{9}{66}, \quad a_5 = \frac{2}{66}, \quad \text{for } t = 2,$$

$$a_1 = \frac{11}{26}, \quad a_2 = \frac{8}{26}, \quad a_3 = \frac{5}{26}, \quad a_4 = \frac{2}{26}, \qquad \text{for } t = 1,$$

and similarly for $t = T$ and $T-1$.

For $c^2 = 1/20$,

$$a_t = \frac{1}{4}, \quad a_{t-1} = a_{t+1} = \frac{3}{16}, \quad a_{t-2} = a_{t+2} = \frac{1}{8}, \quad a_{t-3} = a_{t+3} = \frac{1}{16},$$

$$\text{for } t \geq 3,$$

$$a_1 = \frac{29}{216}, \quad a_2 = \frac{42}{216}, \quad a_3 = \frac{55}{216}, \quad a_4 = \frac{42}{216}, \quad a_5 = \frac{29}{216},$$

$$a_6 = \frac{16}{216}, \quad a_7 = \frac{3}{216}, \qquad \text{for } t = 3,$$

$$a_1 = \frac{40}{185}, \quad a_2 = \frac{51}{185}, \quad a_3 = \frac{40}{185}, \quad a_4 = \frac{29}{185}, \quad a_5 = \frac{18}{185},$$

$$a_6 = \frac{7}{185}, \qquad \text{for } t = 2,$$

$$a_1 = \frac{5}{15}, \quad a_2 = \frac{4}{15}, \quad a_3 = \frac{3}{15}, \quad a_4 = \frac{2}{15}, \quad a_5 = \frac{1}{15}, \quad \text{for } t=1.$$

And so on.

Such technique may be applicable to "smoothing" of time series data, and it has some appealing property when compared with the traditional moving average method in that it has diminishing weights as the distance increases, and that it can also produce the smoothed value for both ends of the series coherently.

It will be desirable to extend this technique to the case under more general assumptions. It will seem to be necessary at least to extend to the case for $h = 1$ of Assumption 2 in §2, i.e. to have predictor or estimator free of the bias due to linear trend. Of course, it is possible to work out for individual case numerically, but unfortunately there is no systematic way of finding a solution in the general cases.

We shall investigate the situation into some detail for $h = 1$. From (2.5) we have

$$f(k) = f(0) + kd + \sum_{j=1}^{k-1} (k-j)\Delta^2 f(j+1), \quad k = 1,2,\ldots, \tag{5.2}$$

where slight and obvious modification of notation is introduced.

Suppose that we would like to esti ate $f(t)$. Then under Assumption 2, for $h = 1$, we must have

$$\begin{aligned} \sum a_k &= 1 \\ \sum k a_k &= t \, . \end{aligned} \tag{5.3}$$

And under this condition, for $\hat{f}(t) = \sum a_k Y_k$,

$$\sup_f \left|E(\hat{f}(t) - f(t)\right| = \sup_\Delta \left| \sum_{j=1}^{t-1} (t-j)\Delta_{j+1}^2 - \sum_{k=1}^{T} \sum_{j=1}^{k-1} (k-j)a_k \Delta_{j+1}^2 \right|$$

$$= \sup_\Delta \left| \sum_{j=1}^{t-1} \{(t-j) - \sum_{k=j+1}^{T} (k-j)a_k\}\Delta_{j+1}^2 \right.$$

$$\left. - \sum_{j=t}^{T-1} \sum_{k=j+1}^{T} (k-j)a_k \Delta_{j+1}^2 \right|$$

$$= \sup_\Delta \left| \sum_{j=1}^{t-1} \left(\sum_{k=1}^{j} (k-j)a_k \right)\Delta_{j+1}^2 - \sum_{j=t}^{T-1} \left(\sum_{k=j+1}^{T} (k-j)a_k \right)\Delta_{j+1}^2 \right|$$

$$= c \left\{ \sum_{j=1}^{t-1} \left| \sum_{k=1}^{j} (j-k)a_k \right| + \sum_{j=t}^{T-1} \left(\sum_{k=j+1}^{T} (k-j)a_k \right| \right\}, \quad (5.4)$$

where $\Delta_{j+1}^2 = \Delta^2 f(j+1)$, for simplicity. We shall denote

$$b_j = \sum_{k=1}^{j} (j-k)a_k \,, \qquad \text{for } j \leq t-1 \,,$$

$$= \sum_{k=j}^{T} (k-j)a_k \,, \qquad \text{for } j \geq t+1 \,,$$

and

$$M = \sum_{j=1}^{t-1} |b_j| + \sum_{j=t+1}^{T} |b_j| \,.$$

Then,

$$\sup E(\hat{f}(t) - f(t))^2 = c^2 M^2 + \sum a_k^2 \,.$$

We have to minimize this value under the constraint (5.3). Similarly with §2, we denote by Φ the Lagrangian form, and we have

$$\frac{1}{2}\left(\frac{\partial \phi}{\partial a_k}\right)_{\pm} = c^2 M \sum_{j=k+1}^{t-1} (j-k)\ \mathrm{sgn}(b_j \pm 0) + a_k - \lambda_0 - k\lambda_1 , \qquad \text{for } k \leq t-1,$$

$$= c^2 M \sum_{j=t+1}^{k-1} (k-j)\ \mathrm{sgn}(b_j \pm 0) + a_k - \lambda_0 - k\lambda_1 , \qquad \text{for } k \geq t+1 ,$$

$$= a_k + \lambda_0 - k\lambda_1 , \qquad \text{for } k = t . \tag{5.4}$$

We can surmize that if t and $T-t$ are both large, $a_k \geq 0$ for all k, and $a_{t+s} = a_{t-s}$, $s = 1, \ldots,$ and $a_{t+s} = a_{t-s} = 0$, for $s \geq m+1$, for some integer m.

If this conjecture were true, we would have $b_j \geq 0$ for all j, and $= 0$, for $j \leq t-m$, and $j \geq t+m$, and we can also suppose that $\lambda_1 = 0$ from the symmetry. Then (5.4) is reduced to

$$\frac{1}{2}\left(\frac{\partial \phi}{\partial a_k}\right)_{\pm} = c^2 M \frac{|t-k|(|t-k|+1)}{2} + a_k - \lambda_0 , \qquad \text{for } |t-k| \leq m-1,$$

$$= \pm c^2 M \frac{|t-k|(|t-k|+1)}{2} - \lambda_0 , \qquad \text{for } |t-k| \geq m .$$

Hence if for

$$a_k = \max \left\{\lambda_1 - \frac{c^2 M}{2} |t-k|(|t-k|+1),\ 0\right\} ,$$

it holds that

$$\sum a_k = 1 \quad \text{and} \quad \sum k\, a_k = t ,$$

and

$$M = \sum_k \frac{|t-k| \ (|t-k|+1)}{2} a_k \ .$$

Then this gives the solution to the problem. The relation between m and c^2 is obtained from

$$\lambda_0 - \frac{c^2M}{2} m(m-1) > 0 \geq \lambda_0 - \frac{c^2M}{2} m(m+1) \ ,$$

and

$$M = 2 \ (\sum_{j=1}^{m-1} \frac{j(j+1)}{2} - \sum_{j=1}^{m-1} \frac{j^2(j+1)^2}{4} c^2M) \ ,$$

and it is shown that

$$\frac{m(m-1)(m-2)(m+1)(2m-1)}{30} < \frac{1}{c^2} \leq \frac{m(m-1)(m+1)(m+2)(2m+1)}{30} \ . (5.5)$$

Consequently, if m determined from (5.5) is not larger than min (t-1, T-t), a_k obtained above gives the solution, thus we have the minimax estimator for intermediate values.

Next we shall consider the case for t = 1. Then it is shown that

$$\frac{1}{2} (\frac{\partial\phi}{\partial a_k})_{\pm} = c^2M \sum_{j=2}^{k-1} (k-j) \ \mathrm{sgn} \ (b_j \pm 0) + a_k - \lambda_0 - k\lambda_1 \quad \text{for } k \geq 2,$$

$$= a_k + \lambda_0 - k\lambda_1 \ , \qquad \text{for } k = 2.$$

And in this case we have

$$\sum k\, a_k = 1 \qquad \text{or} \qquad \sum_{k=2}^{T} (k-1)a_k = 0 .$$

Similar discussion as in §2 for the case $t = 0$, we have also in this case that $b_j \leq 0$, $j = 2,\ldots,$ and

$$\begin{aligned} a_k &= \lambda_0 - k\lambda_1 - \frac{c^2M}{2} k(k-1) , && \text{for } k \leq m , \\ &= 0 , && k = m , \end{aligned}$$

and m is determined so that $a_m < 0$ and

$$\lambda_0 - (m+1)\lambda_1 - \frac{c^2M}{2} m(m+1) \geq 0 .$$

Thus it should be remarked that the coefficients of the minimax predictor have quite different structures for the intermediate and end values, which will be shown by the following illustration (Figure 1).

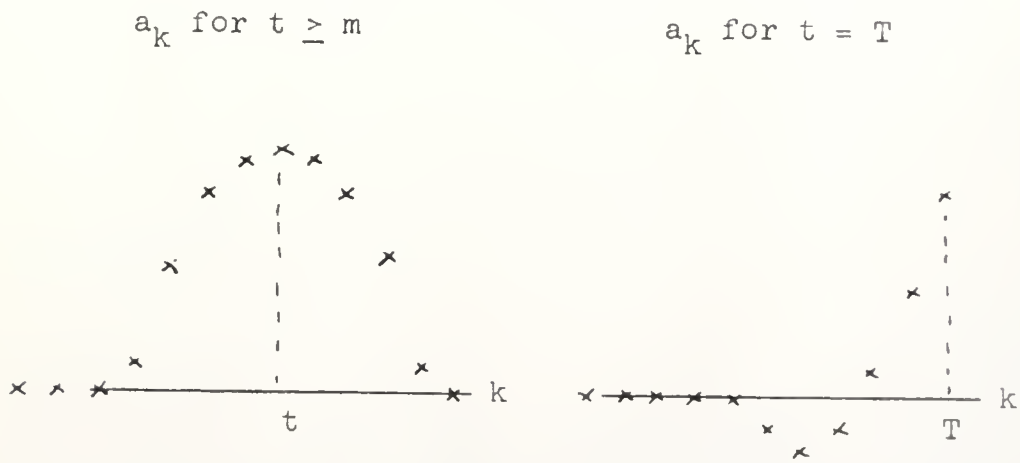

Figure 1

And the case for t near the end is much more complicated and it is difficult to represent a general solution in a systematic way.

And it must also be remarked that the solution in this case is not equal to the weighted least square estimator with weight proportional to $|t - k|$, since for weighted least square estimator a_k must be of the form

$$a_k = |t - k| (\alpha + \beta_k)$$

which is not the case for intermediate k.

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1 ORIGINATING ACTIVITY (Corporate author) Courant Institute of Mathematical Sciences New York University	2a REPORT SECURITY CLASSIFICATION not classified 2b GROUP none

3 REPORT TITLE

Minimax Linear Predictor under Lipschitz' Type Conditions for the Regression Function

4 DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report January 1968

5 AUTHOR(S) (Last name, first name, initial)

Takeuchi, Kei

6 REPORT DATE	7a TOTAL NO OF PAGES	7b NO OF REFS
January 1968	39	none

8a CONTRACT OR GRANT NO. Nonr-285(38) b. PROJECT NO. NR 042-206 c. d.	9a ORIGINATOR'S REPORT NUMBER(S) IMM 363 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) none

10. AVAILABILITY/LIMITATION NOTICES

Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
none	U.S. Navy, Office of Naval Research 207 West 24th St., New York, N.Y.

13. ABSTRACT

Suppose that $Y_i = f(x_i) + U_i$, $i = 1,\ldots,n$, and $E(U_i) = 0$. $E(U_i^2) = \sigma^2$, $E(U_i\, U_j) = 0$, $i \neq j$.

We want to estimate $f(x_0)$ by a linear function $\hat{f}(x_0) = \sum a_i Y_i$. Those coefficients a_i which minimize $\sup_f E(\hat{f}(x_0) - f(x_0))^2$ for some class of f are sought. In two cases where

(a) $\left|\dfrac{f(x_i) - f(x_j)}{x_i - x_j}\right| \leq c\sigma$,

(b) $x_i = i$, $x_0 = 0$

$|\Delta^{h+1} f(k)| \leq c\sigma$ where $\Delta f(k) = f(k+1) - f(k)$,

the minimax solution is obtained, and it is shown that the solution coincides with weight least squares with weight

$$w_i = \max\,(\lambda - \mu|x_i - x_j|,\ 0)\ .$$

DD FORM 1 JAN 64 1473

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through ____________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through ____________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through ____________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.

Head, Logistics and Mathematical Statistics Branch Office of Naval Research Washington, D.C. 20360	3 copies
Commanding Officer Office of Naval Research Branch Office Navy #100 Fleet Post Office New York, New York	2 copies
Defense Documentation Center Cameron Station Alexandria, Virginia 22314	20 copies
Defense Logistics Studies Information Exchange Army Logistics Management Center Fort Lee, Virginia Attn: William B. Whichard	1 copy
Technical Information Officer Naval Research Laboratory Washington, D.C. 20390	6 copies
Commanding Officer Office of Naval Research Branch Office 207 West 24th Street New York, New York 10011 Attn: J. Laderman	1 copy
Commanding Officer Office of Naval Research Branch Office 1030 East Green Street Pasadena 1, California Attn: Dr. A.R. Laufer	1 copy
Bureau of Supplies and Accounts Code OW Department of the Navy Washington 25, D.C.	1 copy
Institute for Defense Analyses Communications Research Division von Neumann Hall Princeton, New Jersey	1 copy
University of Chicago Statistical Research Center Chicago, Illinois Attn: Prof. Paul Meier	1 copy

Stanford University
Applied Mathematics & Statistics Lab.
Stanford, California
Attention: Prof. Gerald J. Lieberman

Florida State University
Department of Statistics
Tallahassee, Florida
Attention: Prof. I. R. Savage

Florida State University
Department of Statistics
Tallahassee, Florida
Attention: Dr. Ralph A. Bradley

Princeton University
Department of Mathematics
Princeton, New Jersey

Columbia University
Department of Mathematical Statistics
New York 27, New York
Attention: Prof. T. W. Anderson

University of California
Department of Statistics
Berkeley 4, California
Attention: Prof. J. Neyman

University of Washington
Department of Mathematics
Seattle 5, Washington
Attention: Prof. Z.W. Birnbaum

Cornell University
Department of Mathematics
Ithaca, New York
Attention: Prof. J. Wolfowitz

Harvard University
Department of Statistics
Cambridge, Massachusetts
Attention: Prof. W. G. Cochran

The Research Triangle Institute
Statistics Research Division
505 West Chapel Hill Street
Durham, North Carolina
Attention: Dr. Malcolm R. Leadbetter

Columbia University
Department of Industrial Engineering
New York 27, New York
Attention: Prof. Cyrus Derma

Columbia University
Department of Mathematics
New York 27, New York
Attention: Prof. H. Robbins

New York University
Courant Institute of Mathematical Sciences
New York, New York 10012
Attention: Prof. W. M. Hirsc

Cornell University
Department of Plant Breeding
Biometrics Unit
Ithaca, New York
Attention: Walter T. Federer

University of North Carolina
Statistics Department
Chapel Hill, North Carolina
Attention: Prof. Walter L. Smith

Stanford University
Department of Statistics
Stanford, California
Attention: Prof. Herman Rubi

Math Department
University of California
San Diego
LaJolla, California 92038

New York University
Department of Industrial Engineering and Operations Research
Bronx 63, New York
Attention: Prof. J. H. Kao

University of Wisconsin
Department of Statistics
Madison, Wisconsin
Attention: Prof. G.E.P. Box

Prof. Anders Hald
Institute of Mathematical Statistics
University of Copenhagen
Copenhagen, Denmark

Yale University
Department of Mathematics
New Haven, Connecticut
Attention: Prof. L. J. Savage

Stanford University
Institute in Engineering and
Economic Systems
School of Engineering
Stanford, California 94305.

Stanford University
Department of Statistics
Stanford, California
Attention: Prof. E. Parzen

Arcon Corporation
Box 419
Wakefield, Mass. 01880
Attention: Dr. Arthur Albert

University of California
Institute of Engineering Research
Berkeley 4, California
Attention: Prof. R. E. Barlow

Department of Probability & Statistics
The University
Sheffield 10, England
Attention: Prof. J. Gani

Dr. Jack R. Borsting
Chairman
Department of Operations Analysis
U. S. Naval Postgraduate School
Monterey, California 93940

Prof. Harold E. Dodge
Rutgers - The State Univ.
Statistics Center
New Brunswick, New Jersey

Yale University
Department of Statistics
New Haven, Connecticut
Attention: Prof. F. J. Anscombe

Purdue University
Division of Mathematical
Sciences
Lafayette, Indiana
Attention: Prof. S.S. Gupta

Cornell University
Department of Industrial
Engineering
Ithaca, New York
Attention: Prof. Robert Bechhofer

Stanford University
Department of Statistics
Stanford, California
Attention: Prof. C. Stein

Applied Mathematics and
Statistics Lab.
Department of Statistics
Stanford University
Stanford, California
Attention: Prof. H. Solomon

Decision Studies Group
460 California Avenue
Palo Alto, California
Attention: Warren R. Ketler

The Johns Hopkins University
Department of Mathematical
Statistics
34th and Charles Streets
Baltimore 18, Maryland
Attention: Prof. Geoffrey S. Watson

DATE DUE	BORROWER'S NAME	ROOM NUMBER

DATE DUE
